

Nama : Fuad Zauqi Nur
Kelas : IF-40-02
NIM : 1301164392

I. Deskripsi Masalah

K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikan berdasarkan mayoritas dari kategori pada KNN. Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. *Classifier* tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah *k* obyek atau (titik *training*) yang paling dekat dengan titik *query*. Klasifikasi menggunakan *voting* terbanyak diantara klasifikasi dari *k* obyek.. algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru.

II. Strategi Penyelesaian Masalah

A. Parameter nilai K

Nilai K yang diambil adalah $K=7$ dengan menggunakan metode, mencoba $K=3$, $K=5$, $K=7$ sehingga pemilihan nilai K berdasarkan hasil yang relevan terhadap data.

B. Perhitungan Jarak antara data Train dan data Test

Logika yang digunakan adalah menguji setiap data Test terhadap 800 data Train yang ada sehingga menghasilkan nilai *Distance* dan terdapat 7 nilai *Distance* pada setiap data Test dikarenakan $K=7$. Perhitungan jarak menggunakan persamaan Manhattan.

C. Pengurutan 7 nilai Jarak minimum yang paling mendekati K

Pengurutan menggunakan algoritma Sorting yaitu Insertion Sort lalu *Looping* saat mencari data yang relevan antara list yang telah di sorting dengan list yang memiliki nilai Y dengan demikian outputan dari bagian ini adalah 7 nilai Jarak minimum yang mendekati K beserta calon kategori pada satu baris pengujian pada Data Test.

D. Voting calon kategori

Voting calon kategori dilakukan agar menentukan hasil akhir nilai Y pada setiap baris uji Data Test. Strategi penyelesaiannya yaitu dengan cara menghitung secara iterasi jumlah dari setiap kategori dengan membandingkan 7 nilai pada baris tersebut dengan list yang terdapat nilai jarak dan nilai Y. Lalu dengan jumlah nilai terbanyak pada suatu kategori akan dinyatakan sebagai nilai Y pada baris tersebut.

Data Output : TebakanTugas3.csv (Terlampir pada folder Data Tugas 3 AI)